Q5

David

2019-03-11

```
library(knitr)
library(twitteR)
consumer_key = "dOLrbYfXjbDo7f7IZ3MFFDF4E"
consumer_secret = "GKbWkDANMboESv0YwvSSZKWivvbIyEzhzZxroDIODY1thrk0Gz"
access_token = "1050995496381120513-8E1z4xtsUYmNTjTRZCRLN4WYMkJjyZ"
access_secret = "bAGtoWn9wR0EUKPybvr7yhK0FFaFJNRviV2LINV06pgRp"
setup_twitter_oauth(consumer_key, consumer_secret,
access_token, access_secret)
```

```
## [1] "Using direct authentication"
```

```
Translink = searchTwitter(searchString =
'Translink', n = 6000, lang = "en")
TL=twListToDF(Translink)
```

a.

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.5.2
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:twitteR':
##
##     id, location
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

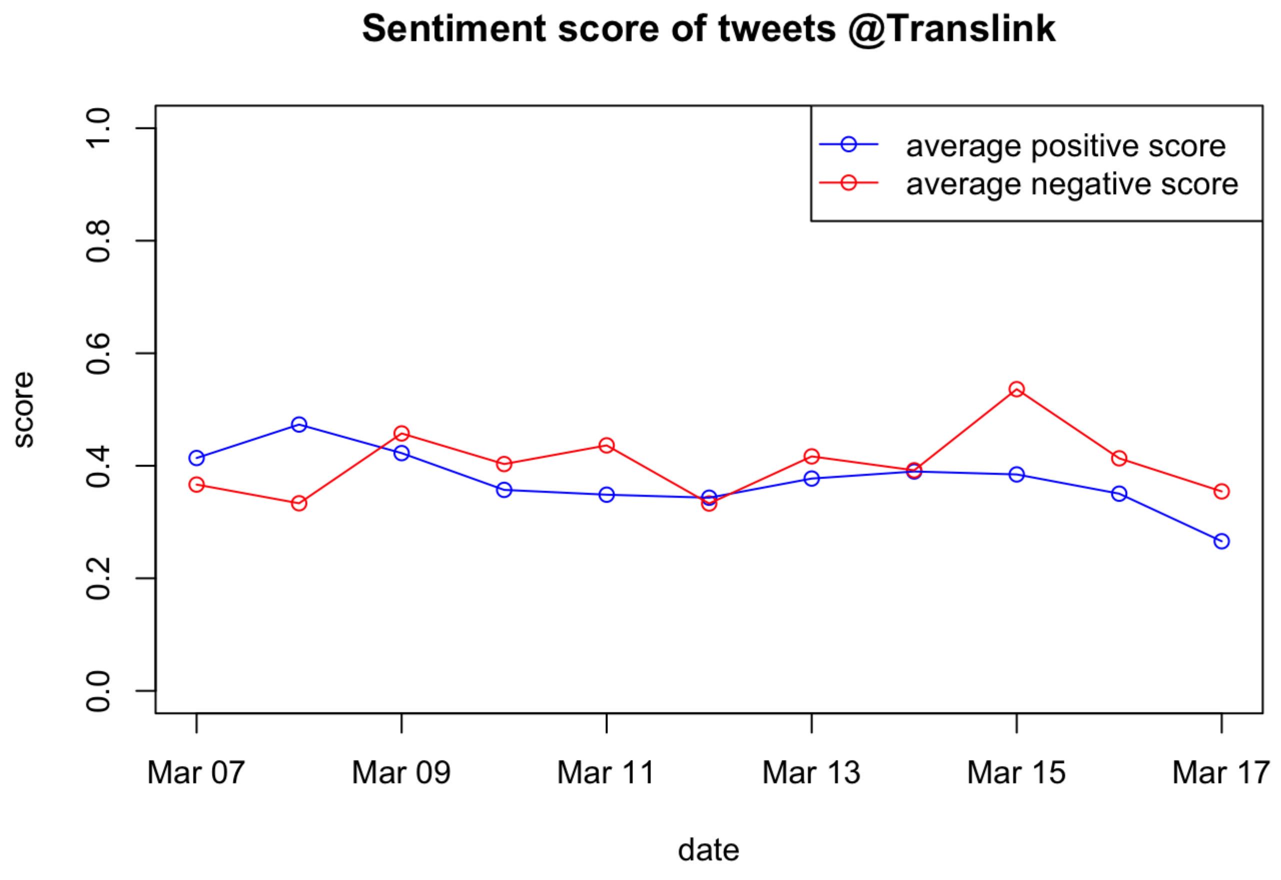
```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
#load files
pos = scan("positive-words.txt", what = "character", comment.char = ";")
neg = scan("negative-words.txt", what = "character", comment.char = ";")
getSentimentScore = function(tweet_text, pos, neg) {
tweet_text = tweet_text$text
tweet_text = gsub("(RT|via)(?:\\b\\W*@\\w+)", " ", tweet_text)
tweet_text = gsub("@\\w+", " ", tweet_text)
tweet_text = gsub("&\\w+;", "", tweet_text)
tweet_text = gsub("[[:cntrl:]]", "", tweet_text)
tweet_text = gsub("(n|N)o.", "number", tweet_text)
tweet_text = gsub("[[:digit:]]", "", tweet_text)
tweet_text = gsub("(?!')[[:punct:]]", "", tweet_text, perl = T)
tweet_text = gsub("(\\B'|'\\B)", "", tweet_text)
tweet_text = gsub("(\\B'|'\\B)", "", tweet_text)
tweet_text = gsub("\\w+\\…","", tweet_text)
tweet_text = iconv(tweet_text, "ASCII", "UTF-8", sub = "")
tweet_text = tolower(tweet_text)
tweet_text = gsub("http\\w+", "", tweet_text)
tweet_text = gsub("[ \t]{2,}", " ", tweet_text)
tweet_text = gsub("^\\s+|\\s+$", "", tweet_text)
  # split into words
  word.list = strsplit(tweet_text, " ")
  # initialize vector to store score
  score = numeric(length(word.list)) # loop through each tweet
  positive = numeric(length(word.list))
  negative = numeric(length(word.list))
  for (i in 1:length(word.list)) {
    # compare our words to the dictionaries of positive # & negative terms
    pos.matches = match(word.list[[i]], pos)
    neg.matches = match(word.list[[i]], neg)
    # match() returns the position of the matched term # or NA we just want a TRUE/FALS
E:
    pos.matches = !is.na(pos.matches)
    neg.matches = !is.na(neg.matches)
    # and conveniently enough, TRUE/FALSE will be # treated as 1/0 by sum():
    score[i] = sum(pos.matches) - sum(neg.matches)
    positive[i] = sum(pos.matches)
    negative[i] = sum(neg.matches)
  }
  return(data.frame(positive_word_count = positive,
                    negative_word_count = negative,
                    sentiment_score = score))
}
SentimentScore=getSentimentScore(TL, pos, neg)
Translink=TL
Translink$created=cut(TL$created, "days")
Translink=cbind(Translink, SentimentScore)
Translink$created=as.Date(Translink$created)
by_days = group_by(Translink, created)
summary_stat = summarise(by_days, Avg.Pos = mean(positive_word_count), Avg.Neg = mean(ne
gative_word_count))
head(summary_stat, 5)
```

```
## # A tibble: 5 x 3
##   created    Avg.Pos Avg.Neg
##   <date>       <dbl>   <dbl>
## 1 2019-03-07   0.414   0.367
## 2 2019-03-08   0.473   0.333
## 3 2019-03-09   0.422   0.458
## 4 2019-03-10   0.357   0.403
## 5 2019-03-11   0.349   0.436
```

b.

```
plot(summary_stat$created, summary_stat$Avg.Pos, col=4, ylim = c(0,1),xlab = "date", yla
b = "score", type = "p", main = "Sentiment score of tweets @Translink")
points(summary_stat$created, summary_stat$Avg.Neg, col=2)
lines(summary_stat$created, summary_stat$Avg.Pos,col=4,lwd=1)
lines(summary_stat$created, summary_stat$Avg.Neg,col=2,lwd=1)
legend(x="topright", pch = c(1,1),lty = c(1,1), col = c(4,2), c("average positive score"
, "average negative score "))
```


Sentiment score of tweets @Translink
average positive score
average negative score
score
Mar 07
Mar 09
Mar 11
Mar 13
Mar 15
Mar 17
date